Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application:

Listing of Claims:

1-6.   (canceled).

7.   (new) A method of encoding data within a system, the method comprising:

determining a set of cut points in input data, the input data including a sequence of symbols, wherein a cut point is determined using a fingerprint representation of a number of sequential symbols in the sequence of symbols;

segmenting the input data as indicated by the set of cut points;

for each segment, determining whether the segment is to be a referenced segment;

for each referenced segment, replacing the segment data of the referenced segment with a reference label;

for each referenced segment not already present in a persistent segment store, storing a reference binding in the persistent segment store, wherein a reference binding associates a referenced segment's data and its reference label;

determining whether any sequence of segments is to be grouped as a reference group;

for each reference group, replacing the references in the group with a group label; and

for each reference group not already present in the persistent segment store, storing a group reference binding in the persistent segment store, wherein a group reference binding associates a reference group's references with its group label.

8.   (new) The method of claim 7, further comprising:

recursively identifying groups of labels into higher level groups, wherein groups of labels are one or more of groups of reference labels and groups of group labels;

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

4   for each higher level group, replacing the higher level group with a group label;

5 and

6   for each higher level group not already present in the persistent segment store,

7 storing a group reference binding in the persistent segment store for the higher level group.

1   9. (new) The method of claim 7, wherein the input data comprises payloads

2 of messages between clients and servers in a client-server network.

1   10. (new) The method of claim 7, wherein the input data comprises portions

2 of files in an on-line backup system, further comprising representing files in the on-line backup

3 system as sequences of at least one of reference labels and group labels, and storing contents of

4 the persistent segment store as part of the on-line backup system.

1   11. (new) The method of claim 7, wherein the input data comprises portions

2 of files in a file system, further comprising representing files in the file system as sequences of at

3 least one of reference labels and group labels and a segment store.

1   12. (new) The method of claim 7, wherein the input data comprises portions

2 of files to be used in a file system, the method further comprising:

3   when storing a file to the file system, encoding it with at least one segment of the

4 file being represented as a segment referenced in the persistent segment store; and

5   when retrieving a file from the file system, caching the file in a local file store as a

6 decoded file, wherein each reference label and each group label is replaced with corresponding

7 segment data from the persistent segment store.

1   13. (new) A method for encoding data in a system, the method comprising:

2   determining a set of cut points for input data based on a fingerprint function, the

3 fingerprint function indicating a cut point based on a number of symbols input into the

4 fingerprint function;

5   segmenting the input data based on the set of cut points;

6   for each segment, determining whether the segment is to be a referenced segment;

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

7   for each referenced segment, replacing segments in the segmented input data with

8   reference labels;

9   for each referenced segment not already present in a persistent segment store,

10   storing a reference binding in the persistent segment store, wherein a reference binding

11   associates a referenced segment's data and its reference label;

12   determining whether a group of reference labels should be grouped as a reference

13   group;

14   for each reference group determined, replacing the references in the group with a

15   group label; and

16   for each reference group not already present in the persistent segment store,

17   storing a group reference binding in the persistent segment store, wherein a group reference

18   binding associates a reference group's references with its group label.

1   14.   (new) The method of claim 13, wherein the fingerprint function

2   comprises a hash function.

1   15.   (new) The method of claim 13, wherein determining the set of cut points

2   comprises:

3   determining a fingerprint window comprising a sequence of input symbols,

4   wherein the fingerprint window is associated with an offset;

5   inputting the sequence of input symbols into the fingerprint function, the

6   fingerprint function outputting a fingerprint value; and

7   determining from the fingerprint value if a cut point should be determined at the

8   offset.

1   16.   (new) The method of claim 15, wherein determining the set of cut points

2   comprises:

3   if it is not determined from the fingerprint value that a cut point should be

4   determined at a new offset, advancing the fingerprint window to comprise a new sequence of

5   input symbols, wherein the fingerprint window is associated with the offset;

6   inputting the new sequence of input symbols into the fingerprint function, the

7   fingerprint function outputting a new fingerprint value; and

8   determining from the new fingerprint value if a cut point should be determined at

9   the new offset.

1   17.   (new) The method of claim 16, further comprising repeating the

2   advancing, inputting, and determining steps until a cut point is determined.

1   18.   (new) The method of claim 13, wherein determining whether the group of

2   references should be grouped as the reference group comprises:

3   inputting the group of references into the fingerprint function, the fingerprint

4   function outputting a fingerprint value; and

5   determining from the fingerprint value if the group of references should be a

6   grouped as a reference group.

1   19.   (new) The method of claim 18, further comprising:

2   if it is not determined from the fingerprint value that should be grouped as the

3   reference group, advancing the fingerprint window to comprise a new group of reference labels;

4   inputting the new group of reference labels into the fingerprint function, the

5   fingerprint function outputting a new fingerprint value; and

6   determining from the new fingerprint value if the new group of reference labels

7   should be a grouped as a reference group.

1   20.   (new) The method of claim 19, further comprising repeating the

2   advancing, inputting, and determining steps until the reference group is determined.

1   21.   (new) The method of claim 13, wherein the reference group comprises at

2   least one of a reference label and input data.

1   22.   (new) The method of claim 13, further comprising sending the segmented

2   input data, the segmented input data including at least one of a reference label and a group label.

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

1   23.   (new) The method of claim 22, further comprising:

2   for each reference label in the segmented input data, retrieving from the persistent

3   segment store the segment's data that is associated with the reference label.

1   24.   (new) The method of claim 22, further comprising:

2   for each group label in the segmented input data, retrieving from the persistent

3   segment store the reference labels that are associated with the group label; and

4   for each reference label retrieved, retrieving from the persistent segment store the

5   segment's data that is associated with the retrieved reference label.

1   25.   (new) An encoder for encoding data, the encoder comprising:

2   an input for receiving input data;

3   fingerprint logic configured to determine a fingerprint representation of a number

4   of sequential symbols in the sequence of symbols;

5   a cutpoint determiner configured to determine a set of cut points in input data,

6   wherein a cut point is determined using the fingerprint representation of the number of sequential

7   symbols in the sequence of symbols;

8   a segmenter configured to segment the input data as indicated by the set of cut

9   points;

10   a replacer comprising:

11   for each segment, logic configured to determine whether the segment is to

12   be a referenced segment;

13   for each referenced segment, logic configured to replace the segment data

14   of the referenced segment with a reference label;

15   for each referenced segment not already present in a persistent segment

16   store, logic configured to store a reference binding in the persistent segment store, wherein a

17   reference binding associates a referenced segment's data and its reference label;

18   logic configured to determine whether any sequence of segments is to be

19   grouped as a reference group;

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

20   for each reference group, logic configured to replace the references in the

21   group with a group label; and

22   for each reference group not already present in the persistent segment

23   store, logic configured to store a group reference binding in the persistent segment store, wherein

24   a group reference binding associates a reference group's references with its group label.

1   26.   (new) The encoder of claim 25, wherein the replacer comprises:

2   logic configured to recursively identify groups of labels into higher level groups,

3   wherein groups of labels are one or more of groups of reference labels and groups of group

4   labels;

5   for each higher level group, logic configured to replace the higher level group

6   with a group label; and

7   for each higher level group not already present in the persistent segment store,

8   logic configured to store a group reference binding in the persistent segment store for the higher

9   level group.

1   27.   (new) The encoder of claim 25, wherein the input data comprises

2   payloads of messages between clients and servers in a client-server network.

1   28.   (new) The encoder of claim 25, wherein the input data comprises portions

2   of files in an on-line backup system, further comprising logic configured to represent files in the

3   on-line backup system as sequences of at least one of reference labels and group labels, and logic

4   configured to store contents of the persistent segment store as part of the on-line backup system.

1   29.   (new) The encoder of claim 25, wherein the input data comprises portions

2   of files in a file system, further comprising logic configured to represent files in the file system

3   as sequences of at least one of reference labels and group labels and a segment store.

1   30.   (new) The encoder of claim 25, wherein the input data comprises portions

2   of files to be used in a file system, the encoder further comprising:

3   logic configured to encode a file with at least one segment of the file being

4   represented as a segment referenced in the persistent segment store when storing the file to the

5   file system; and

6   logic configured to cache a file in a local file store as a decoded file, wherein each

7   reference label and each group label is replaced with corresponding segment data from the

8   persistent segment store when retrieving the file from the file system.

1   31.   (new) A coder for processing data, the coder comprising:

2   a cut point determiner configured to determine a set of cut points for input data

3   based on a fingerprint function, the fingerprint function indicating a cut point based on a number

4   of symbols input into the fingerprint function;

5   a segmenter configured to segment the input data based on the set of cut points;

6   a segment replacer comprising:

7   for each segment, logic configured to determine whether the segment is to

8   be a referenced segment;

9   for each referenced segment, logic configured to replace segments in the

10   segmented input data with reference labels; and

11   for each referenced segment not already present in a persistent segment

12   store, logic configured to store a reference binding in the persistent segment store, wherein a

13   reference binding associates a referenced segment's data and its reference label;

14   a reference replacer comprising:

15   logic configured to determine whether a group of reference labels should

16   be grouped as a reference group;

17   for each reference group determined, logic configured to replace the

18   references in the group with a group label; and

19   for each reference group not already present in the persistent segment

20   store, logic configured to store a group reference binding in the persistent segment store, wherein

21   a group reference binding associates a reference group's references with its group label.

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

1   32.   (new) The coder of claim 31, wherein the fingerprint function comprises a

2   hash function.

1   33.   (new) The coder of claim 31, wherein the cut point determiner is

2   configured to:

3   determine a fingerprint window comprising a sequence of input symbols, wherein

4   the fingerprint window is associated with an offset;

5   input the sequence of input symbols into the fingerprint function, the fingerprint

6   function outputting a fingerprint value; and

7   determine from the fingerprint value if a cut point should be determined at the

8   offset.

1   34.   (new) The coder of claim 33, wherein the cut point determiner is

2   configured to:

3   if it is not determined from the fingerprint value that a cut point should be

4   determined at a new offset, advance the fingerprint window to comprise a new sequence of input

5   symbols, wherein the fingerprint window is associated with the offset;

6   input the new sequence of input symbols into the fingerprint function, the

7   fingerprint function outputting a new fingerprint value; and

8   determine from the new fingerprint value if a cut point should be determined at

9   the new offset.

1   35.   (new) The coder of claim 34, wherein the cutpoint determiner is

2   configured to repeatedly advance, input, and determine until a cut point is determined.

1   36.   (new) The coder of claim 31, wherein the logic configured to determine

2   whether the group of references should be grouped as the reference group comprises:

3   logic to input the group of references into the fingerprint function, the fingerprint

4   function outputting a fingerprint value; and

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

5   logic to determine from the fingerprint value if the group of references should be

6   a grouped as a reference group.

1   37.   (new) The coder of claim 36, wherein the logic configured to determine

2   whether the group of references should be grouped as the reference group comprises:

3   if it is not determined from the fingerprint value that should be grouped as the

4   reference group, logic configured to advance the fingerprint window to comprise a new group of

5   reference labels;

6   logic configured to input the new group of reference labels into the fingerprint

7   function, the fingerprint function outputting a new fingerprint value; and

8   logic configured to determine from the new fingerprint value if the new group of

9   reference labels should be a grouped as a reference group.

1   38.   (new) The coder of claim 37, wherein the reference replacer is further

2   configured to repeatedly advance, input, and determine until the reference group is determined.

1   39.   (new) The coder of claim 31, wherein the reference group comprises at

2   least one of a reference label and input data.

1   40.   (new) The coder of claim 31, further comprising a communicator

2   configured to send the segmented input data, the segmented input data including at least one of a

3   reference label and a group label.

1   41.   (new) The coder of claim 40, further comprising:

2   for each reference label in the segmented input data, a decoder configured to

3   retrieve from the persistent segment store the segment's data that is associated with the reference

4   label.

1   42.   (new) The coder of claim 41, wherein the decoder is configured to:

2   for each group label in the segmented input data, retrieve from the persistent

3   segment store the reference labels that are associated with the group label; and

Appl. No. 10/731,687
Amdt. dated May 27, 2004
Preliminary Amendment

PATENT

4   for each reference label retrieved, retrieve from the persistent segment store the

5   segment's data that is associated with the retrieved reference label.